Submission in Response to NSF CI 2030 Request for Information

DATE AND TIME: 2017-03-06 18:52:49
REFERENCE NO: 178

Author Names & Affiliations

- David Dunigan - University of Nebraska-Lincoln

Contact Email Address (for NSF use only)

(Hidden)

Research Domain, discipline, and sub-discipline

Life Sciences

Title of Submission

Bioinformatics support

Abstract (maximum ~200 words).

In a age of rapidly expanding access to the biosphere via DNA sequencing, we continue to lack appropriate tools and support to evaluate these data. In the sub-discipline of viruses of microbes, the discovery rate of new viruses significantly lags behind the ability to describe environmental sequences that point to novel viruses but cannot be fully characterized at this point. Viruses are the most abundant biological entity of the planet and very likely the most genetically diverse. One bottleneck in the discovery and description of the virosphere is a lack of tools and support personnel for bioinformatics.

Question 1 Research Challenge(s) (maximum ~1200 words): Describe current or emerging science or engineering research challenge(s), providing context in terms of recent research activities and standing questions in the field.

We're trying to describe the virosphere. We know there are a great number of virus-like particles (>10^31 in total; ~10^7 per mL of sea water) found in all corners of the planet. Yet, little is known about them. We are challenged by the lack of trained personnel with skills and interest in virus discovery and characterization from a computational perspective. Also, many of the tools available are "user friendly".

Question 2 Cyberinfrastructure Needed to Address the Research Challenge(s) (maximum ~1200 words): Describe any limitations or absence of existing cyberinfrastructure, and/or specific technical advancements in cyberinfrastructure (e.g. advanced computing, data infrastructure, software infrastructure, applications, networking, cybersecurity), that must be addressed to accomplish the identified research challenge(s).

A primary limitation in virus discovery and characterization is the bottleneck in bioinformatics capacities that include useable tools for extracting information from the ever growing data sets, but also the people trained to work on the cyberinfrastructure side of the problem. This research challenge could be overcome with targeted training programs and research support for the bioinformatics specialists. Too often these individuals are viewed as temporary or "soft money" type positions. Our data sets are not getting smaller. Our needs for bioinformatic interrogation is not going away...it is increasing daily. These people are needed and their positions need to be secure.

Question 3 Other considerations (maximum ~1200 words, optional): Any other relevant aspects, such as organization, process, learning and workforce development, access, and sustainability, that need to be addressed; or any other issues that NSF should consider.

This is concern is mostly at the workforce development level and I believe many domains of the life sciences will benefit with additional resources put toward training of biologists, as well as those with computational skills wanting to move into biology, for tackling the challenging issues before with computer skills.